Chapter 9
Live and Video Simulations of Medical History-Taking: Theoretical Background, Design, Development, and Validation of a Learning Environment

Maximilian C. Fink, Victoria Reitmeier, Matthias Siebeck, Frank Fischer, and Martin R. Fischer

This chapter's simulation at a glance

Domain	Medicine
Topic	Dyspnea diseases occurring in an emergency room setting
Learner's task	Take a full medical history in the role of a physician to diagnose patients with dyspnea
Target group	Advanced medical students and early-career physicians
Diagnostic mode	Individual diagnosing
Sources of information	Information is primarily gathered in interaction with the (video) patient. Some prior information (e.g., laboratory and ECG results) is provided by documents
Special features	The content was created for both live and video simulations

9.1   Introduction

History-taking is an essential diagnostic situation for physicians for two reasons. According to a recent literature review, 60–80% of relevant information in medical diagnosing emerges from history-taking (Keifenheim et al., 2015). Moreover, about

M. C. Fink (✉) · V. Reitmeier · M. Siebeck · M. R. Fischer
Institute for Medical Education, University Hospital, LMU Munich, Munich, Germany
e-mail: Maximilian.Fink@med.uni-muenchen.de

F. Fischer
Chair of Education and Educational Psychology, Department of Psychology, LMU Munich, Munich, Germany

two-thirds of all medical diagnoses can be made accurately after taking a patient's history (Peterson et al., 1992). Even though history-taking is of such great importance, intermediate students still experience difficulties in conducting comprehensive medical interviews for the purpose of diagnosing (Bachmann et al., 2017). Meta-analytic findings indicate that simulation-based learning conveys diagnostic competences effectively if adequate instructional support is offered to learners (Cook et al., 2010, 2013). Instructional support measures such as reflection phases and role-taking seem promising for fostering diagnostic competences in history-taking situations because they are beneficial for acquiring complex skills in other contexts within medical training (Stegmann et al., 2012; Mamede et al., 2008). Presently, however, only limited empirical findings are available concerning facilitating diagnostic competences in history-taking simulations via these two instructional support measures (Keifenheim et al., 2015). Thus, this project aimed firstly to develop realistic history-taking simulations for the assessment of diagnostic competences. In a second step, this project will use these simulations in future studies that vary reflection phases and role-taking. Dyspnea (shortness of breath) was chosen as the key symptom of the cases in the simulations.

9.2 Theoretical Background

9.2.1 Definition and Models of the Medical Interview

The *medical interview* is a dynamic encounter in which a physician and a patient interactively construct the patient's medical history together (Haidet & Paterniti, 2003). This process is called history-taking. History-taking can be supported with assistive resources (e.g., history-taking forms), and takes place in all medical specialties with direct patient contact in diverse care contexts, including emergency medicine, family medicine and psychiatry (Keifenheim et al., 2015). According to popular models of history-taking (Bird & Cohen-Cole, 1990; Smith et al., 2000; Rosenberg et al., 1997; Kurtz et al., 2003), the medical interview can be conceptualized on a continuum from patient-centered to physician-centered. In patient-centered medical interviews, the patient's psychological and social context is explored more extensively and the patient steers parts of the conversation (Henderson et al., 2012). In physician-centered interviews, by contrast, the patient's physical symptoms are in focus and the physician leads the conversation. The medical interview can perform a wide range of communicative functions, including gathering data and making diagnoses, educating and making decisions, and establishing rapport (Roter & Hall, 1987; Jefferson et al., 2013). Depending on the specific context of a medical interview (e.g., an emergency room vs. a routine checkup), a patient-centered or physician-centered approach with the aforementioned communicative functions might be more relevant (Keifenheim et al., 2015). As this project applies an experimental paradigm and focusses on the simulation-based assessment and training of diagnostic competences in emergency room dyspnea cases, a physician-

centered approach emphasizing the functions of gathering data and making diagnoses seems most suitable.

9.2.2 Diagnostic Competences in the Medical Interview

Diagnostic competences have been described on an abstract level using the framework presented in Chap. 2 (Chernikova et al., 2022) and will be specified here in the context of this project.

In all diagnostic settings, diagnostic quality is comprised of diagnostic accuracy and diagnostic efficiency. Diagnostic accuracy generally depends on the correctness of the diagnosis as well as its justification—reasoning for and against the main diagnosis. As it is often not possible to rule out all differential diagnoses in a medical interview without further examinations (Petrusa, 2002), the diagnosis and associated justification may be considered preliminary. Efficiency in history-taking is based not only on time spent, but also on the amount of relevant data gathered in this time and the cost and adverse effects of the examinations and interventions ordered.

The diagnostic process can be operationalized in this context primarily via the diagnostic activities of generating hypotheses, generating and evaluating evidence and drawing conclusions. Hypotheses are frequently formed at the beginning of the medical interview using the patient's background information and initial complaint and are updated over the course of the interview (Pelaccia et al., 2014). Evidence generation takes place in history-taking primarily through asking questions but also includes interpreting visible signs (e.g., paleness as a symptom for pulmonary embolism) and acquiring necessary background information. In the medical interview, evidence evaluation is the analysis of the evidence contained in the background information, the signs and symptoms and the patients' answers. The validity and reliability of the different pieces of evidence can differ significantly and must be determined on a case-by-case basis (Redelmeier et al., 2001). In history-taking, the reliability and validity of evidence can be particularly threatened when information is sensitive or difficult for patients to remember and comprehend. For instance, some patients with extensive medical knowledge present a meticulous documentation of the medication they have taken in the last year in the medical interview, while other patients with low medical knowledge experience difficulties remembering important medication they are currently taking. Drawing conclusions involves weighing the generated and evaluated evidence to make a decision. The result is the creation of a diagnosis and a justification.

According to the theoretical framework presented by this research group, individual prerequisites predict the diagnostic quality and diagnostic process. In the context of the medical interview, the professional knowledge base is a key component of these individual prerequisites (Stark et al., 2011; Förtsch et al., 2018) and can be differentiated into conceptual and strategic knowledge (Schmidmaier et al., 2013; Kopp et al., 2008). Conceptual knowledge is defined as "knowledge about the declarative textbook facts" (Schmidmaier et al., 2013, p. 2), while strategic

knowledge "comprises knowledge about problem-solving strategies and heuristics in the process" (Schmidmaier et al., 2013, p. 2). Both types of knowledge, which form the professional knowledge base relevant for the simulation we present in this chapter, include content on diseases that may cause dyspnea as well as content related to conducting the medical interview.

9.2.3   Simulation-Based Learning and Assessment of History-Taking

We propose that history-taking can be facilitated and assessed with live simulations, video simulations and role-plays. Live simulations employ standardized patients who have been systematically trained to act as patients and display symptoms of diseases authentically (May et al., 2009). Video simulations include interactive videos of patients displaying symptoms. User input can take place through a menu or through free text input that is analyzed automatically, e.g., with natural language processing methods (Cook et al., 2010). In role-plays, students receive a script and play the roles of a physician, patient, and observer according to the script (Joyner & Young, 2006). Each of these simulation modalities has certain advantages and disadvantages in medical training. While live simulations are highly interactive, they require a great deal of administrative effort and produce ongoing high costs. Video simulations are expensive at the time of construction but can then be used indefinitely in digital learning environments without new expenditure. Role-plays are inexpensive but require participants to prepare well before taking part.

As seen in Chap. 2 (Chernikova et al., 2022), theoretical arguments and empirical evidence indicate that simulation-based learning with instructional support is a promising method for facilitating diagnostic competences. With regard to simulation-based learning in history-taking situations, 17 studies had been conducted by the time an extensive literature review appeared in 2015 (Keifenheim et al., 2015). Even though most of these studies reported positive effects of educational interventions, the literature review had limitations. Many of the included studies combined numerous educational interventions (e.g., lectures and small group work), focused on communication skills as an output measure or did not include a performance measure of diagnostic competences in the posttest. Specific results for reflection phases and roles are still not available for this context.

Live simulations have been used to assess performance in medicine for decades (e.g., Harden & Gleeson, 1979), and computer-based simulations have become increasingly popular (Ryall et al., 2016). The literature agrees that simulation-based assessment can be reliable and predict clinical performance (Ryall et al., 2016; Petrusa, 2002; Edelstein et al., 2000). However, the reliability and validity of the assessment depend on factors such as the authenticity and standardization of the simulated situation and patient, the choice of scoring algorithms and determination of expert solutions, and the sampling and number of cases (Petrusa, 2002;

Weller et al., 2005; Clauser & Schuwirth, 2002). In general, it is recommended to use multiple, authentic and well-operationalized cases for assessment and to complement simulation-based assessment with other measures such as knowledge tests (Ryall et al., 2016).

9.2.4 Design, Development and Validation Process Objectives and Research Questions

The project focused in this phase on the creation and validation of live simulations and video simulations as assessment instruments. The main research questions were: Are live and video simulations valid and reliable assessment tools? [RQ 1], Are live and video simulations experienced as authentic? [RQ 2], and Are conceptual and strategic knowledge tests predictive of diagnostic quality? [RQ 3].

9.2.5 Simulation Design and Development

The project team consisted of two professors of medicine with expertise in medical education, one professor of educational psychology, a licensed physician and a Ph. D. student in learning sciences. The physician was mainly responsible for creating the content of the simulations and knowledge tests. The Ph.D. student primarily had the task of designing and conducting the experimental study. The professors acquired funding, supervised the physician and the Ph.D. student and offered feedback and advice on their academic work.

In a first step, dyspnea, the subjective feeling of shortness of breath, was selected as a cardinal symptom because it is one of the most common presentations in emergency rooms and GP practices (Berliner et al., 2016). A blueprint was drafted that specified the diagnoses of three training and six assessment cases. Two of the training cases focused on cardiac insufficiency and pulmonary embolism, while for one training case the diagnosis was COPD. Four of the six assessment cases involved specific types of cardiac insufficiency and pulmonary embolism (near transfer), while for two cases the diagnoses were hyperventilation and pneumonia, which are not similar to any training case (far transfer). Next, a case scenario was created to determine the structure and sequence of all cases. All cases would start with key prior information (such as a pathological laboratory test result or an ECG) and a presentation of the chief complaint by the patient. Then, the cases would proceed and be followed by 8 min of history-taking during which the participant could ask or select questions independently. In this phase, participants would mainly conduct a physician-centered interview, asking or selecting general screening questions (e.g., "Is this the first time you are encountering this problem?") and specific questions to test certain diagnoses (e.g., "Have you had swollen legs?"). The
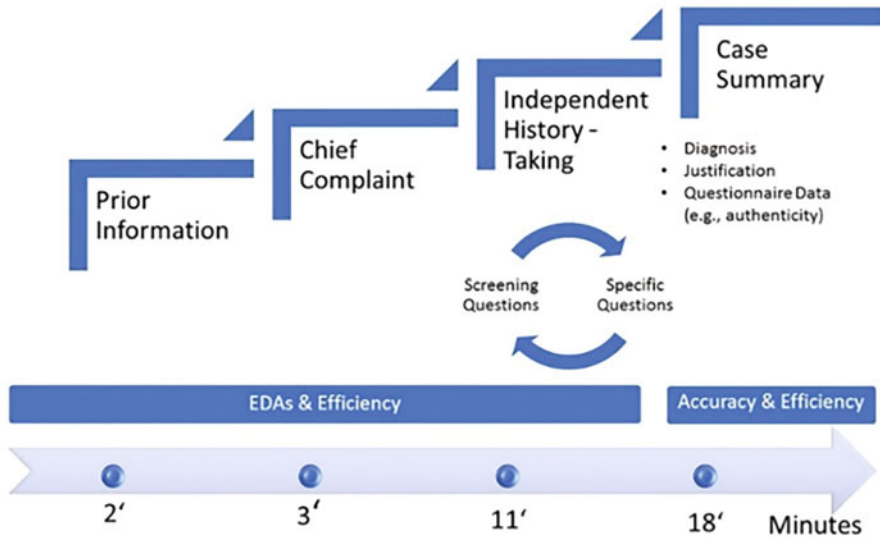
Fig. 9.1 Simulation scenario

questions covered the history-taking categories of principal symptoms, past medical history, allergies and current medication, social and family history and system overview and were based on a classification by Bornemann (2016). Then, students would provide a diagnosis and a justification in a case summary. Figure 9.1 depicts the simulation scenario, including the length of its elements and relevant processes (for more information, see the next section).

Developing a foundation for the live and video simulations, the licensed physician first created a set of history-taking questions as well as nine case vignettes. To create video simulations, a computer scientist programmed a video simulator with a menu and integrated it into the e-learning platform CASUS 3.0 (Instruct, 2018). Professional actors were filmed acting out the cases as standardized patients in a clinical setting and the videos were cut and embedded in the simulator. To produce live simulations, an experimental protocol was created that outlined the behavior of standardized patients and experimenters. The actors were trained to act out the case in face-to-face encounters and individual coaching was offered by the licensed physician. The simulations were conducted in a simulation center at the University Hospital of LMU Munich in Germany that offered three test rooms with a stretcher for the live simulations as well as a computer room for the video simulations and pretest. The final live simulation is displayed in Fig. 9.2 and the final video simulation in Fig. 9.3.
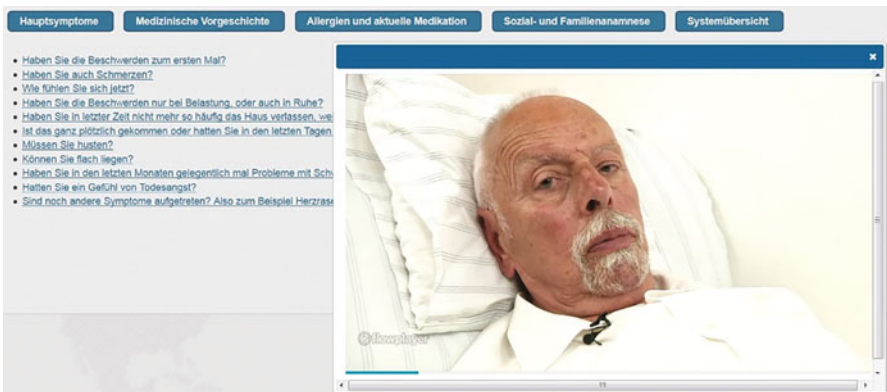
Fig. 9.2 Live Simulation of History-taking

Fig. 9.3 Video Simulation of History-taking

9.2.6   Test Design and Development

To measure diagnostic competences according to the framework described in Chap. 2, separate measures for diagnostic quality, the diagnostic process and the professional knowledge base were created.

Diagnostic quality was assessed with a case summary after each case. Participants listed the final diagnosis in the case summary and provided a justification for this diagnosis. Moreover, participants listed further examinations and treatments. The final diagnosis was chosen from a long menu (i.e., an auto-complete search list that contained many possible diagnoses) and the justifications, examinations and treatments were entered in a text field. Diagnostic accuracy was calculated by adding up partial scores for the final diagnosis (incorrect vs. partially correct vs. correct). The justification for the diagnosis was determined by the percentage of correct reasons mentioned out of all correct reasons for a case defined in the expert solution. Both of

these facets of diagnostic accuracy were coded based on the learners' answers by two physicians with a scoring rubric.

Diagnostic processes were tracked in video simulations with log files and in live simulations with video recordings. Video simulation data was coded automatically using *R* scripts. Live simulation data was coded by trained student assistants with a scoring rubric. In both types of simulations, tracked behaviors and their timestamps facilitated detailed analyses of the diagnostic activities. For instance, we investigated evidence generation in depth by analyzing the number and relevance of questions selected.

To measure the professional knowledge base, a conceptual and strategic knowledge test was created. These knowledge tests were based on the conceptualizations of professional knowledge by Förtsch et al. (2018) and Fischer et al. (2005). The conceptual knowledge test contained 39 questions and covered symptoms, etiology, therapy and further diagnostics and interventions for dyspnea. The questions used were extracted from a professional database for examinations. This knowledge test encompassed multiple-choice questions with a varying number of correct answers. The strategic knowledge test consisted of 10 key feature cases (i.e., short case vignettes that contain crucial clinical problems, see Hrynchak et al., 2014) on the topic of dyspnea that were developed by the physician as part of the project. Each case vignette contained four questions on the diagnosis, history-taking, treatment and further diagnosis.

9.2.7 Cooperation with Other Projects

The materials presented above were developed in collaboration with another project on facilitating cooperative medical diagnosing competences (Radkowitsch et al., 2022). Both projects developed comparable simulation blueprints and used the same case summary. The strategic and conceptual knowledge tests were structured in a similar way. Close collaboration also took place with yet another project on diagnostic competences in the diagnostic interview in mathematics education (Marczynski et al., 2022). This collaboration was mainly related to creating the live simulation. In both projects, similar blueprints were created before writing the case vignettes. Standardized patients and students were trained comparably. Measures of diagnostic accuracy and diagnostic processes were operationalized in a similar way in both projects.

9.2.8 Validation Process

A pilot study, an expert workshop and a validation study were conducted to evaluate topics such as the usability, authenticity and correctness of the simulations and tests and to make revisions.

A sample of $N = 12$ medical students took part in the pilot study. The video simulation in the pilot study involved a prototype of the video simulator programmed by the first author and the live simulation employed trained student assistants as actors. Participants diagnosed one case in the video simulation and one case in the live simulation; the sequence of the simulations was randomized. Initial results of the pilot study showed that participants displayed slightly higher diagnostic accuracy in video simulations than in live simulations, and that live simulations were perceived as more authentic than video simulations (Fink et al., 2018). Because of technical problems with the video simulation, a computer scientist produced a professionally programmed video simulation for the validation study. It also became evident that non-professional actors in the live simulation did not act in a highly standardized and authentic way. Therefore, professional actors with experience as standardized patients rather than student assistants were trained to act in the live simulation for the validation study.

To evaluate the authenticity and difficulty of the nine developed case vignettes, an expert workshop with $N = 8$ licensed physicians was conducted. The physicians judged seven case vignettes as authentic and of adequate difficulty for the study and suggested major revisions to two cases. Modifications were made accordingly before all scripts for the video simulation were filmed and before actors prepared for the live simulation.

A total of $N = 86$ medical students took part in the validation study. The study used a mixed design with the between factor expertise (novices vs. interns) and the within factor sequence (video simulations—live simulations vs. live simulations—video simulations). Participants were eligible if they were either in the first 2 years of the clinical part of their university education (novices) or in their final clinical year (interns). Moreover, participants had to be fluent in German to rule out possible effects of language competence. The study used the final live and video simulations presented in this chapter. Participants were randomly assigned to one of the two sequences and took part in a pretest of conceptual and strategic knowledge and then solved three cases in each sequence. Initial findings indicate higher diagnostic accuracy of student participants in live than in video simulations (Fink et al., 2019). These findings are opposed to the findings of the pilot study. Due to the revised simulations, the larger sample, and the higher number of cases, the results of the validation study seem more reliable. Moreover, similarly to the pilot study, live simulations were perceived as more authentic than video simulations. The created knowledge tests were reliable and differentiated between novices and interns. In correlational analyses of the validity of the different knowledge tests and simulations, strategic and conceptual knowledge correlated positively with diagnostic performance in the simulations. Both types of knowledge correlated positively with each other.

All in all, the reported findings demonstrate that live simulations are suitable for the reliable and valid assessment of diagnostic competences in history-taking and offer even higher interactivity and authenticity than video simulations. The created video simulations may still require certain changes, such as longer familiarization with the history-taking menu, to achieve comparable validity and reliability to live simulations and may then be suitable for the economical and standardized

assessment of medical interviewing skills. The validity and reliability of the developed knowledge tests were confirmed.

9.2.9 Conclusion Summary

This chapter reported on the theoretical background and the design, development, and validation process of a research project investigating the facilitation of diagnostic competences in live and video history-taking simulations.

In the section on the theoretical background, the summarized models of history-taking showed that a physician-centered approach to history-taking that emphasizes the functions of gathering data and making diagnoses is suitable for the assessment of diagnostic competences in experimental settings. Moreover, the section on diagnostic competences in the medical interview adapted the conceptual model presented in Chap. 2 to history-taking by presenting sensible operationalizations of diagnostic accuracy, delineating the major diagnostic activities (i.e., generating hypotheses, generating and evaluating evidence and drawing conclusions), and specifying the topics relevant for the assessment of professional knowledge in this situation. In addition, possible benefits and drawbacks of live simulations, video simulations and role-plays were outlined. The summary of key findings on training and assessing history-taking with simulations demonstrated that the differential effects of role-taking and reflection phases need further research.

The section on the design, development, and validation process highlighted the importance of systematic design, expert workshops, pilot studies and validation studies. It contains materials and operationalizations for future studies and programs seeking to design interactive history-taking simulations. The presented materials also show how comparable live and video simulations can be designed and developed. Findings from the validation study suggest that the created simulations may be employed after making minor changes, and the knowledge tests assess separate but related aspects of diagnostic competences validly and reliably.

9.2.10 Open Questions for Research

In line with Question 2 of the overarching research questions mentioned in the introduction by Fischer and Opitz (2022), future studies within this project will investigate the effect of instructional support measures on the acquisition of diagnostic competences. More precisely, the project will examine the effect of reflection phases and role-taking in live and video simulations and role-plays. Even though reflection phases have been shown to be effective instructional support measures (Mamede et al., 2008, 2012; Mann et al., 2009), it is currently not clear whether reflection in video simulations during problem-solving or after problem-solving is more effective and what learning mechanisms, such as the application of

certain types of professional knowledge, make reflection phases effective. Another open research question that also contributes to Question 2 of the overarching research questions described in the introduction by Fischer and Opitz (2022) pertains to the effect of roles in live simulations. Learners in live history-taking simulations can learn to take on the roles of a physician, a patient and an observer. While it has been shown that learning in the agent role is effective, there is a scarcity of findings on the patient and observer role (Cook, 2014). As also pointed out for reflection phases, it must be investigated what learning mechanisms arise in different roles. Finally, the effects of roles and reflection phases should also be explored in role-plays. Only a few findings on this topic are available, and these results do not directly relate to diagnostic competences but typically to communication skills (e.g., Lane & Rollnick, 2007).

The project also plans to contribute new findings to the overarching research question 4 mentioned in the introduction by Fischer and Opitz (2022), which addresses how simulations can be adapted to fit learners. We believe an especially interesting question concerns how adaptive scaffolding could facilitate diagnostic competences in video history-taking simulations. One interesting type of adaptive scaffolding to investigate would be the individual selection of cases of suitable typicality for learners. Case typicality denotes the degree to which a certain case corresponds with the prototypical signs and symptoms of a diagnosis (Papa, 2016). Learners could benefit from adapted case typicality by learning on optimally challenging cases in their zone of proximal development, scaffolded by instructional support (Vygotsky, 1978). Another interesting type of scaffolding to examine would be the adaptive use of reflection phases or examples. It is currently not clear whether the meta-analytical finding that examples are more beneficial for novices than reflection phases and reflection phases are more beneficial for advanced learners than examples (Chernikova et al., 2019) can be replicated in an experimental setting. Furthermore, it is unknown how reflection phases and examples interact in simulation-based learning from atypical cases.

References

Bachmann, C., Roschlaub, S., Harendza, S., Keim, R., & Scherer, M. (2017). Medical students' communication skills in clinical education: Results from a cohort study. *Patient Education and Counseling, 100*, 1874–1881. https://doi.org/10.1016/j.pec.2017.05.030

Berliner, D., Schneider, N., Welte, T., & Bauersachs, J. (2016). The differential diagnosis of dyspnea. *Deutsches Ärzteblatt International, 113*, 834–845. https://doi.org/10.3238/arztebl.2016.0834

Bird, J., & Cohen-Cole, S. A. (1990). The three-function model of the medical interview. In M. S. Hale (Ed.), *Methods in teaching consultation-liaison psychiatry. Advances in psychosomatic medicine* (Vol. 20, pp. 65–88). Karger.

Bornemann, B. (2016). *Dokumentationsbögen der Inneren Medizin und der Chirurgie für Anamnese und körperliche Untersuchung für die studentische Lehre in Deutschland* [documentation forms of internal medicine and surgery for history taking and the physical examination for the medical training of students in Germany: An analysis of content and structure]. Diss. Institut für Didaktik und Ausbildungsforschung in der Medizin der Ludwig-Maximilians-Universität.

Chernikova, O., Heitzmann, N., Fink, M. C., Timothy, V., Seidel, T., & Fischer, F. (2019). Facilitating diagnostic competences in higher education: A meta-analysis in medical and teacher education. *Educational Psychology Review, 32*, 157–196. https://doi.org/10.1007/s10648-019-09492-2

Chernikova, O., Heitzmann, N., Opitz, A., Seidel, T., & Fischer, F. (2022). A theoretical framework for fostering diagnostic competences with simulations. In F. Fischer & A. Opitz (Eds.), *Learning to diagnose with simulations—Examples from teacher education and medical education*. Springer.

Clauser, B. E., & Schuwirth, L. W. T. (2002). The use of computers in assessment. In G. R. Norman, C. P. M. van der Vleuten, & D. I. Newble (Eds.), *International handbook of research in medical education* (pp. 757–792). Springer.

Cook, D. A. (2014). How much evidence does it take? A cumulative meta-analysis of outcomes of simulation-based education. *Medical Education, 48*, 750–760. https://doi.org/10.1111/medu.12473

Cook, D. A., Erwin, P. J., & Triola, M. M. (2010). Computerized virtual patients in health professions education: A systematic review and meta-analysis. *Academic Medicine, 85*, 1589–1602. https://doi.org/10.1097/ACM.0b013e3181edfe13

Cook, D. A., Hamstra, S. J., Brydges, R., Zendejas, B., Szostek, J. H., Wang, A. T., et al. (2013). Comparative effectiveness of instructional design features in simulation-based education: Systematic review and meta-analysis. *Medical Teacher, 35*, e867–e898. https://doi.org/10.3109/0142159X.2012.714886

Edelstein, R. A., Reid, H. M., Usatine, R., & Wilkes, M. S. (2000). A comparative study of measures to evaluate medical students' performances. *Academic Medicine, 75*, 825–833. https://doi.org/10.1097/00001888-200008000-00016

Fink, M. C., Fischer, F., Siebeck, M., Gerstenkorn, H., & Fischer, M. R. (2018). Diagnoseakkuratheit und Authentizität in live- und Videosimulationen von Anamnesegesprächen: Ergebnisse einer Pilotstudie [diagnostic accuracy and authenticity of live and video simulations of history taking: Results of a pilot study]. In *Paper presented at the Jahrestagung der Gesellschaft für Medizinische Ausbildung (GMA), Wien, September, 19*.

Fink, M. C., Siebeck, M., Fischer, F., & Fischer, M. R. (2019). Assessing diagnostic competencies with standardized patients and interactive video simulations: Results from a study on history taking. In *Paper presented at the RIME 2019, Copenhagen, May 24*.

Fischer, M. R., Kopp, V., Holzer, M., Ruderich, F., & Junger, J. (2005). A modified electronic key feature examination for undergraduate medical students: Validation threats and opportunities. *Medical Teacher, 27*, 450–455. https://doi.org/10.1080/01421590500078471

Förtsch, C., Sommerhoff, D., Fischer, F., Fischer, M. R., Girwidz, R., Obersteiner, A., et al. (2018). Systematizing professional knowledge of medical doctors and teachers: Development of an interdisciplinary framework in the context of diagnostic competences. *Educational Sciences, 8*, 207. https://doi.org/10.3390/educsci8040207

Haidet, P., & Paterniti, D. A. (2003). "Building" a history rather than "taking" one: A perspective on information sharing during the medical interview. *Archives of Internal Medicine, 163*, 1134–1140. https://doi.org/10.1001/archinte.163.10.1134

Harden, R. M., & Gleeson, F. A. (1979). Assessment of clinical competence using an objective structured clinical examination (OSCE). *Medical Education, 13*, 39–54. https://doi.org/10.1111/j.1365-2923.1979.tb00918.x

Henderson, M. R., Tierney, L. M., & Smetana, G. W. (Eds.). (2012). *The patient history: An evidence-based approach to differential diagnosis* (2nd ed.). McGraw-Hill.

Hrynchak, P., Glover Takahashi, S., & Nayer, M. (2014). Key-feature questions for assessment of clinical reasoning: A literature review. *Medical Education, 48*, 870–883. https://doi.org/10.1111/medu.12509

Instruct. (2018). *Casus [Computer software]: E-teaching and e-learning software for virtual patients*. https://www.instruct.eu/casus/virtual-patients-software

Jefferson, L., Bloor, K., Birks, Y., Hewitt, C., & Bland, M. (2013). Effect of physicians' gender on communication and consultation length: A systematic review and meta-analysis. *Journal of Health Services Research & Policy, 18*, 242–248. https://doi.org/10.1177/1355819613486465

Joyner, B., & Young, L. (2006). Teaching medical students using role play: Twelve tips for successful role plays. *Medical Teacher, 28*, 225–229. https://doi.org/10.1080/01421590600711252

Keifenheim, K. E., Teufel, M., Ip, J., Speiser, N., Leehr, E. J., Zipfel, S., et al. (2015). Teaching history taking to medical students: A systematic review. *BMC Medical Education, 15*, 159. https://doi.org/10.1186/s12909-015-0443-x

Kopp, V., Stark, R., & Fischer, M. R. (2008). Fostering diagnostic knowledge through computer-supported, case-based worked examples: Effects of erroneous examples and feedback. *Medical Education, 42*, 823–829. https://doi.org/10.1111/j.1365-2923.2008.03122.x

Kurtz, S., Silverman, J., Benson, J., & Draper, J. (2003). Marrying content and process in clinical method teaching: Enhancing the Calgary-Cambridge guides. *Academic Medicine, 78*(8), 802–809.

Lane, C., & Rollnick, S. (2007). The use of simulated patients and role-play in communication skills training: A review of the literature to august 2005. *Patient Education and Counseling, 67*, 13–20. https://doi.org/10.1016/j.pec.2007.02.011

Mamede, S., Schmidt, H. G., & Penaforte, J. C. (2008). Effects of reflective practice on the accuracy of medical diagnoses. *Medical Education, 42*, 468–475. https://doi.org/10.1111/j.1365-2923.2008.03030.x

Mamede, S., van Gog, T., Moura, A. S., de Faria, R. M. D., Peixoto, J. M., Rikers, R. M. J. P., et al. (2012). Reflection as a strategy to foster medical students' acquisition of diagnostic competence. *Medical Education, 46*, 464–472. https://doi.org/10.1111/j.1365-2923.2012.04217.x

Mann, K., Gordon, J., & MacLeod, A. (2009). Reflection and reflective practice in health professions education: A systematic review. *Advances in Health Sciences Education, 14*, 595–621. https://doi.org/10.1007/s10459-007-9090-2

Marczynski, B., Kaltefleiter, L. J., Siebeck, M., Wecker, C., Stürmer, K., & Ufer, S. (2022). Diagnosing 6h Graders' understanding of decimal fractions—fostering mathematics pre-service Teachers' diagnostic competences with simulated one-to-one interviews. In F. Fischer & A. Opitz (Eds.), *Learning to diagnose with simulations—Examples from teacher education and medical education*. Springer.

May, W., Park, J. H., & Lee, J. P. (2009). A ten-year review of the literature on the use of standardized patients in teaching and learning: 1996–2005. *Medical Teacher, 31*, 487–492. https://doi.org/10.1080/01421590802530898

Papa, F. J. (2016). A dual processing theory based approach to instruction and assessment of diagnostic competencies. *Medical Science Educator, 26*, 787–795. https://doi.org/10.1007/s40670-016-0326-8

Pelaccia, T., Tardif, J., Triby, E., Ammirati, C., Bertrand, C., Dory, V., et al. (2014). How and when do expert emergency physicians generate and evaluate diagnostic hypotheses? A qualitative study using head-mounted video cued-recall interviews. *Annals of Emergency Medicine, 64*, 575–585. https://doi.org/10.1016/j.annemergmed.2014.05.003

Peterson, M. C., Holbrook, J. H., Hales, D., & Smith, N. L. (1992). Contributions of the history, physical examination, and laboratory investigation in making medical diagnoses. *Western Journal of Medicine, 156*(2), 163–165.

Petrusa, E. R. (2002). Clinical performance assessments. In G. R. Norman, C. P. M. van der Vleuten, & D. I. Newble (Eds.), *International handbook of research in medical education* (pp. 673–709). Springer.

Radkowitsch, A., Sailer, M., Fischer, M. R., Schmidmaier, R., & Fischer, F. (2022). Learning collaborative diagnosing in medical education—Diagnosing a patient's disease in collaboration with a simulated radiologist. In F. Fischer & A. Opitz (Eds.), *Learning to diagnose with simulations—Examples from teacher education and medical education*. Springer.

Redelmeier, D. A., Tu, J. V., Schull, M. J., Ferris, L. E., & Hux, J. E. (2001). Problems for clinical judgement: 2. Obtaining a reliable past medical history. *Canadian Medical Association Journal, 164*(6), 809–813.

Rosenberg, E. E., Lussier, M.-T., & Beaudoin, C. (1997). Lessons for clinicians from physician-patient communication literature. *Journal of the American Medical Association, 6*(3), 279–283.

Roter, D. L., & Hall, J. A. (1987). Physicians' interviewing styles and medical information obtained from patients. *Journal of General Internal Medicine, 2*, 325–329. https://doi.org/10.1007/BF02596168

Ryall, T., Judd, B. K., & Gordon, C. J. (2016). Simulation-based assessments in health professional education: A systematic review. *Journal of Multidisciplinary Healthcare, 9*, 69–82. https://doi.org/10.2147/JMDH.S92695

Schmidmaier, R., Eiber, S., Ebersbach, R., Schiller, M., Hege, I., Holzer, M., et al. (2013). Learning the facts in medical school is not enough: Which factors predict successful application of procedural knowledge in a laboratory setting? *BMC Medical Education, 13*, 722. https://doi.org/10.1186/1472-6920-13-28

Smith, R. C., Marshall-Dorsey, A. A., Osborn, G. G., Shebroe, V., Lyles, J. S., Stoffelmayr, B. E., et al. (2000). Evidence-based guidelines for teaching patient-centered interviewing. *Patient Education and Counseling, 39*(1), 27–36.

Stark, R., Kopp, V., & Fischer, M. R. (2011). Case-based learning with worked examples in complex domains: Two experimental studies in undergraduate medical education. *Learning and Instruction, 21*, 22–33. https://doi.org/10.1016/j.learninstruc.2009.10.001

Stegmann, K., Pilz, F., Siebeck, M., & Fischer, F. (2012). Vicarious learning during simulations: Is it more effective than hands-on training? *Medical Education, 46*, 1001–1008. https://doi.org/10.1111/j.1365-2923.2012.04344.x

Vygotsky, L. (1978). Interaction between learning and development. In M. Gauvin & M. Cole (Eds.), *Readings on the development of children* (pp. 34–40). Scientific American Books.

Weller, J. M., Robinson, B. J., Jolly, B., Watterson, L. M., Joseph, M., Bajenov, S., et al. (2005). Psychometric characteristics of simulation-based assessment in anaesthesia and accuracy of self-assessed scores. *Anaesthesia, 60*, 245–250. https://doi.org/10.1111/j.1365-2044.2004.04073.x